Accurate sensing of multiple ligands with a single receptor

Vijay Singh[1] and Ilya Nemenman[1,2]

[1]*Department of Physics, Emory University, Atlanta, GA 30322, USA*
[2]*Department of Biology, Emory University, Atlanta, GA 30322, USA*

Cells use surface receptors to estimate the concentration of external ligands. Limits on the accuracy of such estimations have been well studied for pairs of ligand and receptor species. However, the environment typically contains many ligands, which can bind to the same receptors with different affinities, resulting in cross-talk. In traditional rate models, such cross-talk prevents accurate inference of individual ligand concentrations. In contrast, here we show that knowing the precise timing sequence of stochastic binding and unbinding events allows one receptor to provide information about multiple ligands simultaneously and with a high accuracy. We argue that such high-accuracy estimation of multiple concentrations can be realized by the familiar kinetic proofreading mechanism.

Introduction: Cells obtain information about their environment by capturing ligand molecules with receptors on their surface and estimating the ligand concentration from the receptor activity. Limits on the accuracy of such estimation have been a subject of interest since the seminal work of Berg and Purcell [1], with several substantial extensions found recently [2–8]. All of these assume one ligand species coupled to one receptor species. However, cells carry many types of receptors and have many species of ligands around them. The same ligands can bind to many receptors, albeit with different affinities, and vice versa. This is commonly referred to as *cross-talk*.

In traditional deterministic chemical kinetics, one cannot estimate concentrations of more ligands than there are receptor types. Further, even a weak cross-talk prevents determination of concentrations of individual chemical species since activity of a receptor is a function of a weighted sum of concentrations of all ligands that can bind to it. In contrast, here we argue that, with cross-talk, concentration of more than one chemical species can be inferred from the activity of one receptor, provided that the entire stochastic temporal sequence of receptor binding and unbinding events is accessible instead of its mean occupancy. This surprising result can be understood by noting that a typical duration of time that a ligand remains bound to the receptors depends on its unbinding rate. Thus observing the statistics of the receptor's unbound time durations allows estimation of a weighted average of all chemical species that interact with it [5], and then observing the statistics of the bound time durations allows to tell how common each ligand is.

In this article, we derive these results for the simplest problem of the class, namely one receptor interacting with two ligand species. While the exact solution of the inference problem for finding both ligand concentrations is hard to implement using common biochemical machinery, we show that an accurate approximation is possible using the familiar kinetic proofreading mechanism [9, 10].

The Model: Consider a single receptor estimating concentrations of a cognate and a non-cognate ligand, Fig. 1. The ligands bind to the receptor with on-rates $k_{\rm c}$ and $k_{\rm nc}$. These are proportional to the ligand concentrations with known coefficients of proportionality. Thus estimating $k_{\rm c,nc}$ is equivalent to estimating the concentrations themselves. The unbinding, or off-rates, $r_{\rm c}$ and $r_{\rm nc}$, distinguish the two ligands: $r_{\rm nc} > r_{\rm c}$, and a cognate molecule typically stays bound for longer. Following Ref. [5], we estimate $k_{\rm c}$ and $k_{\rm nc}$ from the time-series of binding, $\{t_i^{\rm b}\}$, and unbinding, $\{t_i^{\rm u}\}$ events of a total duration T using Maximum Likelihood techniques. The numbers of binding and unbinding events are different by, at most, one, which is insignificant since we consider $T \to \infty$. Thus without loss of generality, we assume that the first event was a binding event at $t_1^{\rm b}$, and the last one was the unbinding at $t_n^{\rm u}$. We write the probability distribution of observing the sequence $\{t_1^{\rm b}, t_1^{\rm u}, \ldots, t_n^{\rm b}, t_n^{\rm u}\}$, or alternatively the sequence of binding and unbinding intervals $\tau_i^{\rm b} = t_i^{\rm u} - t_i^{\rm b}$, and $\tau_i^{\rm u} = t_{i+1}^{\rm b} - t_i^{\rm u}$:

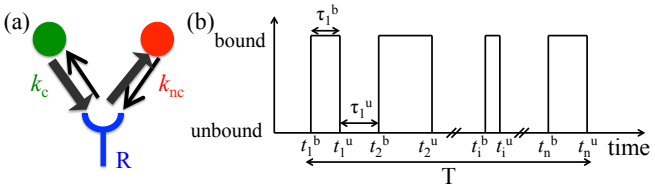

FIG. 1: **The model.** (a). Two lignads, cognate and non-cognate, bind to a receptor R with binding rates $k_{\rm c}$ and $k_{\rm nc}$, respectively. The cognate unbinding rate is defined as lower than the non-cognate one ($r_{\rm c} < r_{\rm nc}$). (b) Time series of receptor occupancy is used to determine both on-rates.

$$P \equiv P(\{\tau_i^{\rm b},\ \tau_i^{\rm u}\}|k_{\rm c}, k_{\rm nc}) = \frac{1}{Z}\prod_{i=1}^{n}\Big[e^{-\tau_i^{\rm u}(k_{\rm c}+k_{\rm nc})} \times \Big(k_{\rm c}\, r_{\rm c}\, e^{-\tau_i^{\rm b} r_{\rm c}} + k_{\rm nc}\, r_{\rm nc}\, e^{-\tau_i^{\rm b} r_{\rm nc}}\Big)\Big]. \quad (1)$$

Here the first term under the product sign is the probability of the receptor staying unbound for $\tau_i^{\rm u}$. The second term, which we from now on denote by $D(k_{\rm c}, k_{\rm nc}, \tau_i^{\rm b})$, is proportional to the probability of staying bound for $\tau_i^{\rm b}$, which has contributions from being bound to the cognate and the noncognate ligands, with odds of $k_{\rm c}/k_{\rm nc}$.

Finally, Z is the normalization. Note that here we define $\tau_n^{\rm u} = t_1^{\rm b} + (T - t_n^{\rm u})$, so that the n'th unbound interval includes the "incomplete" unbound intervals before the first binding and after the last unbinding.

The log-likelihood of $k_{\rm c,nc}$ is the logarithm of P, Eq. (1). Taking the derivatives of the log-likelihood w. r. t. $k_{\rm c}$ and $k_{\rm nc}$ and setting them to zero gives the Maximum Likelihood (ML) equations for the two concentrations. Denoting by $T^{\rm u} = \sum_{i=1}^n \tau_i^{\rm u}$ the total time the receptor is unbound, these are

$$-T^{\rm u} + \sum_{i=1}^{n} \frac{r_{\rm c} e^{-\tau_i^{\rm b} r_{\rm c}}}{D(k_{\rm c}^*, k_{\rm nc}^*, \tau_i^{\rm b})} = 0, \tag{2}$$

$$-T^{\rm u} + \sum_{i=1}^{n} \frac{r_{\rm nc} e^{-\tau_i^{\rm b} r_{\rm nc}}}{D(k_{\rm c}^*, k_{\rm nc}^*, \tau_i^{\rm b})} = 0, \tag{3}$$

where $*$ denotes the ML solution. Multiplying Eqs. (2, 3) by $k_{\rm c}^*$ and $k_{\rm nc}^*$, respectively, and adding them gives

$$k_{\rm c}^* + k_{\rm nc}^* = \frac{n}{T^{\rm u}}, \tag{4}$$

which determines the sum of the two concentrations, showing that the estimates are negatively correlated. As in Ref. [5], the total on-rate (the weighted average of the external concentrations) is determined only by the average duration of the unbound interval, $(n/T^{\rm u})^{-1}$, because no binding is possible when the receptor is already bound.

In general, the ML equations cannot be solved analytically, requiring numerical approaches. However, as all ML estimators, they are unbiased to the leading order in n. The standard errors of the ML estimates can be obtained by inverting the Hessian matrix,

$$\left.\frac{\partial^2 \log P}{\partial k_{\cdot} \partial k_{\cdot}}\right|_{k_{\rm c}^*, k_{\rm nc}^*} = \sum_{i=1}^{n} \left[\frac{-1}{D(k_{\rm c}, k_{\rm nc}, \tau_i^{\rm b})^2} \times \begin{pmatrix} r_{\rm c}^2 e^{-2\tau_i^{\rm b} r_{\rm c}} & r_{\rm c} r_{\rm nc} e^{-\tau_i^{\rm b}(r_{\rm c}+r_{\rm nc})} \\ r_{\rm c} r_{\rm nc} e^{-\tau_i^{\rm b}(r_{\rm c}+r_{\rm nc})} & r_{\rm nc}^2 e^{-2\tau_i^{\rm b} r_{\rm nc}} \end{pmatrix} \right], \tag{5}$$

where $\cdot$ stands for $\{\rm c, nc\}$. The inverse of $\frac{\partial^2 \log P}{\partial k_{\cdot} \partial k_{\cdot}}$, which scales as $\propto 1/n$, sets the minimum variance of any unbiased estimator according to the Cramer-Rao bound. It has straightforward analytical approximations in various regimes. For example, for $k_{\rm c}/k_{\rm nc} \gg 1$ and $r_{\rm c}/r_{\rm nc} \ll 1$, when the noncognate ligand is almost absent, and its few molecules do not bind for long, one gets $\sigma^2(k_{\rm c}^*) \approx \left(\partial^2 \log P/\partial k_{\rm c}^2\right)^{-1}_{k_{\rm c}=k_{\rm c}^*} \approx 1/n$, matching the accuracy of sensing one ligand with one receptor [5]. A regime relevant for detection of a rare, but highly specific ligand [11, 12]) can be investigated as well. Instead, we focus on how the receptor estimates (rather than detects) concentrations of *both* ligands simultaneously, which requires us to investigate the full range of on-rates.

To study the variability of the ML estimator, we define its error as $E_{\rm c,nc} = n\sigma^2(k_{\rm c,nc}^*)/k_{\rm c,nc}^2$, the squared coefficient of variation, multiplied by n, which has a finite limit at $n \to \infty$. $E = 1$ corresponds to the accuracy that a receptor measuring a single ligand would obtain [5]. We show $\log_{10} E$ for different on- and off-rates in Fig. (2). If the two ligands are readily distinguishable, $r_{\rm c} \ll r_{\rm nc}$, then the ligand with the dominant k has $E \sim 1$. When $k_{\rm c} \sim k_{\rm nc}$, $E. \sim 4 \ldots 5$, and it grows to $10 \ldots 30$ for a ligand with a very small relative on-rate. Emphasizing the importance of the time scale separation, $E > 100$ if the ligands are hard to distinguish, $r_{\rm c} \sim r_{\rm nc}$. Here, in addition, the correlation coefficient ρ of the two estimates reaches -1 because the same binding event can be attributed to either ligand. Finally, the asymmetry of the plots w. r. t. the exchange of $k_{\rm c}$ and $k_{\rm nc}$ is because the cognate ligand can generate short binding events, while long events from the noncognate ligand are exponentially unlikely. In summary, it is possible to infer two ligand concentrations from one receptor, with the error of only $1 \ldots 10$ times larger than for ligand-receptor pairs with no cross talk, as long as the two off-rates are substantially different.

Approximate solution. Solving Eqs. (2, 3) to find the ML on-rates would be hard for the cell. Luckily, an approximate solution exists. To find it, we notice that most of the long binding events come from the cognate ligand since the noncognate one dissociates faster. Defining long events as $\tau_i^{\rm b} \geq T^{\rm c}$, we rewrite Eqs. (2, 4) as

$$\frac{n}{k_{\rm c}^* + k_{\rm nc}^*} = \left(\sum_{\tau_i^{\rm b} \geq T^{\rm c}} + \sum_{\tau_i^{\rm b} < T^{\rm c}} \right) \frac{r_{\rm c} e^{-\tau_i^{\rm b} r_{\rm c}}}{D(k_{\rm c}^*, k_{\rm nc}^*, \tau_i^{\rm b})} \tag{6}$$

Assuming that almost all long events are cognate, $T^{\rm c} \gg 1/r_{\rm nc}$, this gives

$$\frac{n}{k_{\rm c}^{\rm a} + k_{\rm nc}^{\rm a}} = \frac{n_{\rm l}}{k_{\rm c}^{\rm a}} + \sum_{\tau_i^{\rm b} < T^{\rm c}} \frac{r_{\rm c} e^{-\tau_i^{\rm b} r_{\rm c}}}{D(k_{\rm c}^{\rm a}, k_{\rm nc}^{\rm a}, \tau_i^{\rm b})}, \tag{7}$$

where $n_{\rm l}$ is the number of long events, and the superscript "a" stands for the *a*pproximate solution. If further T is long enough so that there are many short events, and a single binding duration hardly affects $k_{\rm c}^*$, then the sum in Eq. (7) can be approximated by the expectation value:

$$\frac{n}{k_{\rm c}^{\rm a} + k_{\rm nc}^{\rm a}} = \frac{n_{\rm l}}{k_{\rm c}^{\rm a}} + (n - n_l) \int_0^{T_{\rm c}} \frac{r_{\rm c} e^{-\tau^{\rm b} r_{\rm c}} P(\tau^{\rm b} | k_{\rm c}^{\rm a}, k_{\rm nc}^{\rm a}) d\tau^{\rm b}}{D(k_{\rm c}^{\rm a}, k_{\rm nc}^{\rm a}, \tau^{\rm b})}, \tag{8}$$

where $P(\tau^{\rm b} | k_{\rm c}^{\rm a}, k_{\rm nc}^{\rm a})$ is the probability of observing a binding event of the duration $\tau^{\rm b}$ for the given binding rates,

$$P(\tau^{\rm b} | k_{\rm c}^{\rm a}, k_{\rm nc}^{\rm a}) = \frac{D(k_{\rm c}^{\rm a}, k_{\rm nc}^{\rm a}, \tau^{\rm b})}{k_{\rm c}^{\rm a} + k_{\rm nc}^{\rm a}}. \tag{9}$$

Plugging Eq. (9) into Eq. (8), we obtain

$$\frac{1}{k_{\rm c}^{\rm a} + k_{\rm nc}^{\rm a}} = \frac{n_{\rm l}}{n k_{\rm c}^{\rm a}} + \left(1 - \frac{n_{\rm l}}{n}\right) \frac{1 - e^{-r_{\rm c} T^{\rm c}}}{k_{\rm c}^{\rm a} + k_{\rm nc}^{\rm a}}. \tag{10}$$

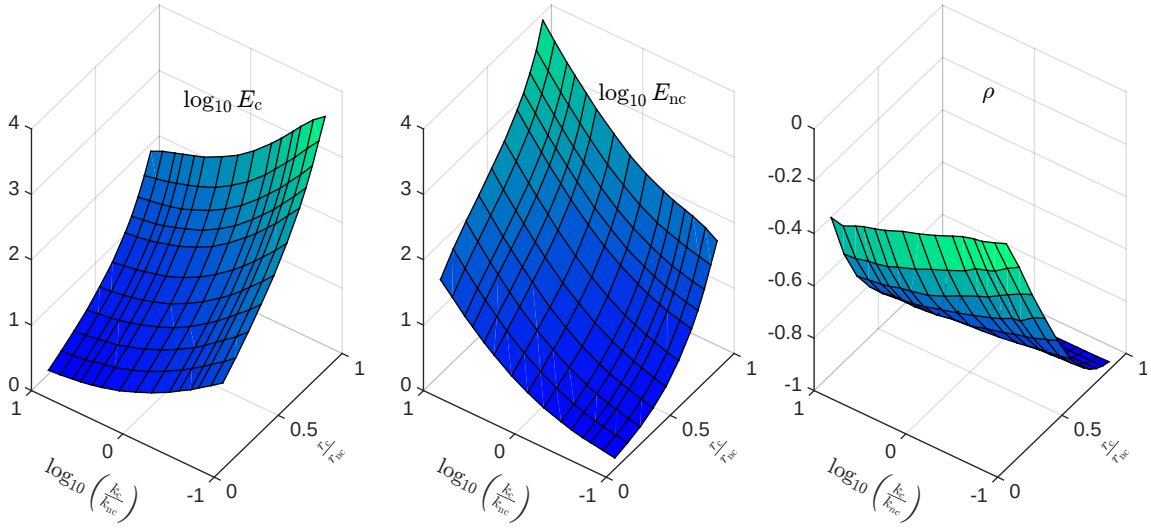

FIG. 2: **Variability of the ML estimators**, represented by $\log_{10} E_c$ (left), $\log_{10} E_{nc}$ (center), and the correlation coefficient ρ between k_c^* and k_{nc}^* (right) as functions of $k_\cdot$ and $r_\cdot$. Here we use $r_{nc} = k_c + k_{nc} = 1$. The plotted quantities are estimated as averages over 30,000 randomly generated binding/unbinding sequences for each combination of the rates. Each sequence consists of $n = 30,000$ binding events, simulated using the Gillespie algorithm. Standard errors are too small to be represented.

Finally, since $n_l \ll n$, using Eq. (4), we get:

$$k_c^a = \frac{n_l}{T^u} e^{r_c T^c}, \tag{11}$$

$$k_{nc}^a = \frac{n}{T^u} - \frac{n_l}{T^u} e^{r_c T^c}. \tag{12}$$

In other words, the approximate cognate ligand concentration is proportional to the number of long events.

We can estimate the bias and the variance of $k_{c,nc}^a$ in a limiting case. If r_c and r_{nc} are not very different from each other, then T^c must be much larger than the inverse of either of them, $T^c \gg \{r_{nc}^{-1}, r_c^{-1}\}$, and $n_l \ll n$. Then most of the variance of $k_{c,nc}^a$ in Eqs. (11, 12) comes from variability of n_l, but not T^u. Thus we write $\langle k_c^a \rangle \approx \frac{\langle n_l \rangle}{\langle T^u \rangle} e^{r_c T^c}$. Further, the individual unbound periods are independent, so that $\langle T^u \rangle = n \langle \tau^u \rangle = n/(k_c + k_{nc})$ (notice the use of k rather than k^a in this relation). Further, $\langle n_l \rangle = n\, P(\tau^b > T^c) = \frac{n}{k_c + k_{nc}} \left(k_c e^{-r_c T^c} + k_{nc} e^{-r_{nc} T^c} \right)$. Combining these expressions, we get

$$\langle k_c^a \rangle \approx k_c + k_{nc} e^{-(r_{nc} - r_c) T^c}. \tag{13}$$

Thus for large T^c, the bias of the approximate estimator, $k_{nc} e^{-(r_{nc} - r_c) T^c}$, grows with the relative number of noncognate long bindings events. In turn, the latter is proportional to k_{nc}, but decreases exponentially with T^c.

Within the same approximation, the variance of the estimator is $\sigma^2(k_c^a) \approx \frac{\sigma^2(n_l)}{\langle T^u \rangle^2} e^{2 r_c T^c}$. But long binding events are rare, independent of each other, and hence obey the Poisson statistics. Thus $\sigma^2(n_l) = \langle n_l \rangle$, so that

$$\sigma^2(k_c^a) \approx \langle k_c^a \rangle \frac{k_c + k_{nc}}{n} e^{r_c T^c}. \tag{14}$$

The variance obviously grows with T^c.

Knowing that the bias and the variance of the approximation change in opposite directions with T^c, we can find the optimal cutoff by minimizing the overall error, or, in other words, solving the bias-variance tradeoff:

$$T_*^c = \arg\min_{T^c} L = \arg\min_{T^c} \left[(k_c - \langle k_c^a \rangle)^2 + \sigma^2(k_c^a) \right], \tag{15}$$

where L is the sum of the squared bias and the variance of the estimator. Near the optimal cutoff, the bias is small, and we use k_c instead of k_c^a for the variance of the estimator, Eq. (14). Then solving Eq. (15) gives:

$$T_*^c = \frac{1}{(2 r_{nc} - r_c)} \log \left[2 T^u \left(\frac{r_{nc}}{r_c} - 1 \right) \frac{k_{nc}^2}{k_c} \right]. \tag{16}$$

Plugging this into Eqs. (13, 14), we can get the minimal error of the estimator, which we omit here for brevity.

The optimal cutoff is $\propto 1/r_{nc}$ if $r_{nc} \gg r_c$, and it grows with r_c, allowing for better disambiguation of cognate and noncognate events. Crucially, the off-rates are specified with the ligand identities. In contrast, the on-rates, $k_{c,nc}$, are what the receptors measures. Therefore, it is encouraging that T^c depends only logarithmically on the on-rates (and also on the duration of the measurement, T^u): fixing T^c as T_*^c at some fixed values of $k_{c,nc}$ remains near-optimal for a broad range of on-rates. To illustrate this, we use $T^c = T_*^c(k_c = k_{nc} = 1/2) \equiv T_0$ and analyze the quality of the approximation in Fig. 3, where we plot the ratio $L_{c,nc}(T_0)/\sigma^2_{k_{c,nc}}$. Since the ratio approaches 1 when $r_c/r_{nc} \to 0$ (specifically, for $r_c/r_{nc} = 0.1$, $L_c(T_0)/\sigma^2_{k_c} \approx 1.47$, and $L_{nc}(T_0)/\sigma^2_{k_{nc}} \approx 1.21$), we conclude that the approximation is accurate even at

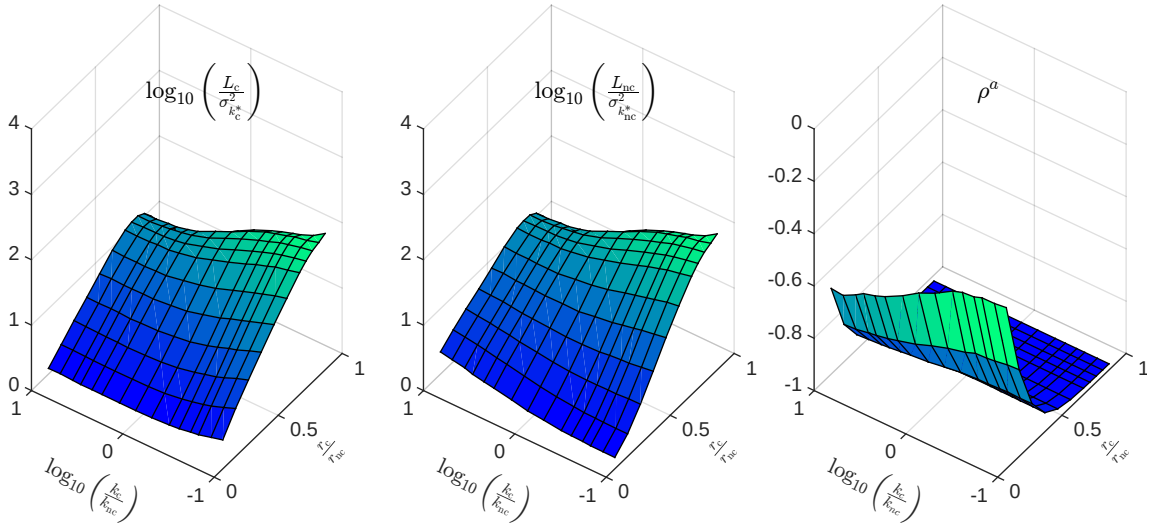

FIG. 3: **Comparison of errors of the approximate and the ML solutions**. We plot $\log_{10}(L_c(T_0)/\sigma^2_{k_c^*})$ (left), $\log_{10}(L_{nc}(T_0)/\sigma^2_{k_{nc}^*})$ (center) and the covariance of the approximate estimates (right) as functions of on- and off-rates. Simulations are performed in the same way as in Fig. 2.

fixed $T^c = T_0$ when its assumptions are satisfied. In contrast, when the ligands are nearly indistinguishable, $L_{c,nc}(T_0)/\sigma^2_{k_{c,nc}} \sim 100$, but here one would not use one receptor to estimate two concentrations since even the ML solution is bad (cf. Fig. 2). Note also that both L_c and L_{nc} are smaller for $r_c \sim r_{nc}$ if $k_c \gg k_{nc}$. This is because our main assumption (that almost all long events are cognate) holds better when cognate ligands dominate. Finally, the correlation coefficient between the approximate estimates, ρ^a (right panel) reaches -1 earlier than in Fig. 2. This is a direct consequence of Eqs. (11, 12).

Kinetic Proofreading for approximate estimation. The approximate solution can be computed by cells using the well-known kinetic proofreading (KPR) mechanism [9, 10, 13, 14]. In the simplest model of KPR [15], intermediate states between an inactive and an active state of a receptor delay the activation. Thus bound ligands can dissociate before the receptor activates, at which point it quickly reverts to the inactive state. Since $r_c > r_{nc}$, cognate ligands dominate among bindings that actually lead to activation. The resulting increase in specificity in various KPR schemes has led to their exploration in the context of *detection* of rare ligands [11, 12, 14], and here we extend them to *measurement* of concentration of cognate and noncognate ligands simultaneously.

Consider a biochemical network in Fig. 4: the receptor (R) activates two messenger molecules (A) and (B). The first one is activated with the rate k_A whenever the receptor is bound. The second one is activated only if the receptor stays bound for longer than a certain T^c (with the delay achieved using the KPR intermediate states). The activation rate after the delay is k_B. The molecules deactivate with the rates r_A and r_B, respectively, and all

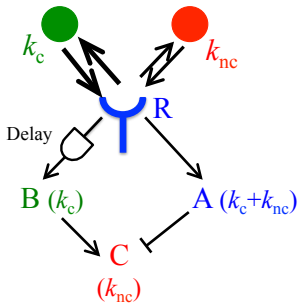

FIG. 4: **Kinetic Proofreading for estimating multiple concentrations.** Molecules A and B are produced when the receptor is bound, but A is produced only for long bindings. Another chemical species C subtracts A from B, so that A approximates k_c and C approxiates k_{nc}.

activations/deactivations are first-order reactions. Then the mean concentrations of the messenger molecules are:

$$\bar{A} = \frac{k_c/r_c + k_{nc}/r_{nc}}{1 + k_c/r_c + k_{nc}/r_{nc}} \frac{k_A}{r_A} \tag{17}$$

$$\bar{B} = \frac{k_c/r_c e^{-r_c T^c} + k_{nc}/r_{nc} e^{-r_{nc} T^c}}{1 + k_c/r_c + k_{nc}/r_{nc}} \frac{k_B}{r_B}, \tag{18}$$

Assuming again that most bindings longer than T^c are cognate, we solve Eqs. (17, 18) for the on-rates

$$k_c = \frac{\bar{B} e^{r_c T^c} r_c r_B}{k_B} \left(1 + \frac{\bar{A}}{k_A/r_A - \bar{A}}\right), \tag{19}$$

$$k_{nc} = \left[\frac{\bar{A}}{k_A/r_A - \bar{A}} - \frac{\bar{B} e^{r_c T^c} r_B}{k_B} \left(1 + \frac{\bar{A}}{k_A/r_A - \bar{A}}\right)\right] r_{nc}. \tag{20}$$

The corrections of the form $\bar{A}/(k_\mathrm{A}/r_\mathrm{A}-\bar{A})$ appear because bindings only happen to unbound receptors, as emphasized in Ref. [5]. However, these nonlinear relations are still hard to implement with simple biochemical components. We solve this by further assuming $\epsilon = \bar{A}/(k_\mathrm{A}/r_\mathrm{A}) \ll 1$, which is true if the receptor is mostly unbound (both on-rates are small compared to the respective off-rates). This gives

$$k_\mathrm{c}^\mathrm{KPR} \approx \frac{\bar{B}e^{r_\mathrm{c}T^\mathrm{c}}r_\mathrm{c}r_\mathrm{B}}{k_\mathrm{B}}, \tag{21}$$

$$k_\mathrm{nc}^\mathrm{KPR} \approx \left(\frac{r_\mathrm{A}\bar{A}}{k_\mathrm{A}} - \frac{\bar{B}e^{r_\mathrm{c}T^\mathrm{c}}r_\mathrm{B}}{k_\mathrm{B}}\right) r_\mathrm{nc}. \tag{22}$$

These equations are analogous to Eqs. (11, 12). They are easy to realize biochemically (cf. Fig. 4): k_c is related to the concentration of the proofread species B by a rescaling, and k_nc comes from subtracting rescaled versions of A and B from each other. The subtraction can be done by the third species C, activated by A and suppressed by B. Since $\epsilon \ll 1$, then $\bar{A}$ and $\bar{B}$ are small, and many such activation-suppression schemes are linearized as the subtraction [8].

The bias of $k_\mathrm{c,nc}^\mathrm{a}$ due to long, but noncognate binding events, Eq. (13), carries over to $k_\mathrm{c,nc}^\mathrm{KPR}$. However, there is an additional contribution since the time to traverse the intermediate states is random. Thus T^c has some variance $\sigma^2_{T^\mathrm{c}}$ [15, 16]. This variability changes the rate of occurence of long biding events, but they are still rare, nearly independent, and Poisson-distributed. Denoting by $\langle\cdot\rangle$ the averaging at a fixed T^c, and by $\bar{\cdot}$ the averaging over T^c, we get

$$\frac{\overline{\langle n_\mathrm{l}\rangle}}{k_\mathrm{c}} \approx \frac{n}{k_\mathrm{c}+k_\mathrm{nc}} e^{-r_\mathrm{c}\bar{T}^\mathrm{c}+\frac{1}{2}r_\mathrm{c}^2\sigma^2_{T^\mathrm{c}}}. \tag{23}$$

Thus $\sigma^2_{T^\mathrm{c}}$ effectively renormalizes the cutoff to $\bar{T}^\mathrm{c} - \frac{1}{2}r_\mathrm{c}\sigma^2_{T^\mathrm{c}}$, which is independent of the on-rates. Replacing T^c in Eqs. (21, 22) by its renormalized value, which is an easy change in the scaling factors, removes this additional bias due to the random T^c in the KPR scheme.

Since long bindings are rare, the variance of the KPR estimator is dominated again generally by $\bar{B}$, but not $\bar{A}$. The intrinsic stochasticity in production of molecules of B contributes to the variance. However, this contribution can be made arbitrarily small by increasing k_B, and we neglect it here. A larger contribution comes from the random number of long bound intervals and a random duration of each of them. To calculate this, in the limit of rare long binding events, we use well-known results in the theory of noise propagation in chemical networks [17]

$$\frac{\sigma_\mathrm{B}^2}{\bar{B}^2} \approx \frac{(1+k_\mathrm{c}/r_\mathrm{c}+k_\mathrm{nc}/r_\mathrm{nc})\, e^{r_\mathrm{c}T^\mathrm{c}-\frac{1}{2}r_\mathrm{c}^2\sigma^2_{T^\mathrm{c}}}}{k_\mathrm{c}(1/r_\mathrm{c}+1/r_\mathrm{B})} = \frac{e^{r_\mathrm{c}T^\mathrm{c}-\frac{1}{2}r_\mathrm{c}^2\sigma^2_{T^\mathrm{c}}}}{k_\mathrm{c}(1/r_\mathrm{c}+1/r_\mathrm{B})} + O(\epsilon). \tag{24}$$

This is a direct analog of Eq. (14).

Discussion. The realization of Refs. [5, 18] and others that the detailed temporal sequence of binding and unbinding events carries more information about the ligand concentration than the mean receptor occupancy is a conceptual breakthrough. It parallels the realization in the computational neuroscience community that precise timing of spikes carries more information about the stimulus than the mean neural firing rate [19–24], and it has a potential to be equally impactful. This extra information when measuring one ligand concentration with one receptor [5] amounted to increasing the sensing accuracy by a constant prefactor, or, equivalently, getting only a finite number of additional bits from even a very long measurement [25]. In contrast, here we show that two concentrations can be measured with one receptor with the variance that decreases inversely proportionally to the number of observations, n, Eq. (14), or to the integration time, $1/r_\mathrm{B}$, Eq. (24), so that the accuracy is only a (small) prefactor lower than would be possible with one receptor per ligand species. Asymptotically, this doubles the information obtained by the receptor [25].

In principle, one can measure more than two concentrations similarly, as long as all species have sufficiently distinct off-rates. While the error (the variance for the ML estimator, and both the bias and the variance for the approximate and the KPR estimators) would grow with a larger number of ligand species, this would still represent a dramatic increase in the information gained by the receptor that keeps track of its precise temporal dynamics, rather than just the average binding state.

Crucially, such improvement would not be possible without the cross-talk, or binding among noncognate ligands and receptors. Normally, the cross-talk is considered a nuisance that must be suppressed [26, 27]. Instead we argue that cross-talk can be beneficial by recruiting more receptor types to measure concentration of the same ligand. In particular, this allows having fewer receptor than ligand species, potentially illuminating how cells function reliably in chemically complex environments with few receptor types. Further, the cross-talk can increase the dynamic range of the entire system: a ligand may saturate its cognate receptor, preventing accurate measurement of its (high) concentration, but it may be in the sensitive range of non-cognate receptors at the same time. Finally, the increased bandwidth may lead to improvements in sensing a time-dependent ligand concentration [11, 18]. We will explore such many-to-many sensory schemes, extending ideas of Ref. [28] to tracking temporal sequences of activation of receptor and to varying environments in forthcoming publications.

While the exact maximum likelihood inference of multiple concentrations from a temporal binding-unbinding sequence is rather complex, we showed that when the cognate and the non-cognate off-rates are substantially different, there is a simpler, approximate, but accurate

inference procedure. In various immune system problems, $r_{\rm nc}/r_{\rm c} \sim 5$, which would allow the approximation to work. Moreover, when the receptor is not saturated and spends most of its time unbound, this inference can be performed by biochemical motifs readily available to the cell. Namely, one needs two branches of activation downstream of the receptor, with one of them having a kinetic proofreading (KPR) time delay, and then an estimate of the difference of activities of the branches. This suggests a possible signal estimation role for the KPR scheme in addition to the more traditional signal detection one [11, 12, 18]. Such branching and merging of signaling pathways downstreams of a receptor is common in signaling [27, 29]. Thus exploring the function of such complex organization in the context of estimation of multiple signals with cross-talk is in order.

In summary, monitoring precise temporal sequences of receptor activation/deactivation opens up new and exciting possibilities for environment sensing by cells.

Acknowledgements: This work was supported in part by the James S. McDonnell foundation and the National Science Foundation.

[1] H. Berg and E. Purcell, Biophys J **20**, 193 (1977).
[2] D. Bray, M. D. Levin, and C. J. Morton-Firth, Nature **393**, 85 (1998).
[3] W. Bialek and S. Setayeshgar, Proc Natl Acad Sci (USA) **102**, 10040 (2005).
[4] R. Endres and N. Wingreen, Proc Natl Acad Sci (USA) **105**, 15749 (2008).
[5] R. G. Endres and N. S. Wingreen, Phys Rev Lett **103**, 158101 (2009).
[6] B. Hu, W. Chen, W.-J. Rappel, and H. Levine, Physi Rev Lett **105**, 048104 (2010).
[7] K. Kaizu, W. de Ronde, J. Paijmans, K. Takahashi, F. Tostevin, and P. R. ten Wolde, Biophys J **106**, 976 (2014).
[8] A. Mugler, A. Levchenko, and I. Nemenman, arXiv:1505.04346 (2015).
[9] J. J. Hopfield, Proc Natl Acad Sci (USA) **71**, 4135 (1974).
[10] J. Ninio, Biochimie **57**, 587 (1975).
[11] J.-B. Lalanne and P. Francois, Proc Natl Acad Sci (USA) **112**, 1898 (2015).
[12] P. Francois, G. Voisinne, E. Siggia, G. Altan-Bonnet, and M. Vergassola, Proc Natl Acad Sci (USA) **110**, E888 (2013).
[13] T. McKeithan, Proc Natl Acad Sci (USA) **92**, 5042 (1995).
[14] B. Goldstein, D. Coombs, J. Faeder, and W. Hlavacek, Adv Exp Med Biol **640**, 82 (2008).
[15] G. Bel, B. Munsky, and I. Nemenman, Phys Biol **7**, 0610003 (2010).
[16] X. Cheng, L. Merchan, M. Tchernookov, and I. Nemenman, Phys Biol **10**, 035008 (2013).
[17] J. Paulsson, Phys Life Rev pp. 157–175 (2005).
[18] E. D. Siggia and M. Vergassola, Proceedings of the National Academy of Sciences **110**, E3704 (2013).
[19] S. Strong, R. Koberle, R. R. d. R. v. Steveninck, and W. Bialek, Phys Rev Lett **80**, 197 (1998).
[20] P. Reinagel and R. C. Reid, J Neurosci **20**, 5392 (2000).
[21] R. Liu, S. Tzonev, S. Rebrik, and K. Miller, J Neurophysiol **86**, 2789 (2001).
[22] I. Nemenman, G. D. Lewen, W. Bialek, and R. R. de Ruyter van Steveninck, PLoS Comp Biol **4**, e1000025 (2008).
[23] A. Fairhall, E. Shea-Brown, and A. Barreiro, Curr Opin Neurobiol **22**, 653 (2012).
[24] C. Tang, D. Chehayeb, K. Srivastava, I. Nemenman, and S. J. Sober, PLoS Biol **12**, e1002018 (2014).
[25] W. Bialek, I. Nemenman, and N. Tishby, Neural Comput **13**, 2409 (2001).
[26] M. McClean, A. Mody, J. Broach, and S. Ramanathan, Nat Gen **39**, 409 (2007).
[27] M. Laub and M. Goulian, Ann Rev Genet **41**, 121 (2007).
[28] J. Tsitron, A. Ault, J. Broach, and A. Morozov, PLoS Comput Biol **7**, e1002224 (2011).
[29] R. Cheong, A. Rhee, C. Wang, I. Nemenman, and A. Levchenko, Science **334**, 354 (2011).